



Data skills in schools: Teaching Statistics

David Young^{1,2}, John Reilly³

¹University of Strathclyde, Glasgow, UK

² NHS Greater Glasgow and Clyde, Glasgow, UK

³Education Scotland, Glasgow, UK

Abstract

In Scotland, numeracy across learning has been emphasised as a key area of the Curriculum for Excellence to provide learners with essential analytic, problem-solving and decision-making skills. The Statistics Award was developed in collaboration with the Scottish Qualifications Authority and the University of Strathclyde in an attempt to provide an opportunity for young people to learn these skills. The emphasis of the course is on the application of statistics to real-life data and interpretation of results in an appropriate context. Minitab or R Studio are used for the analysis of data with learning skills focussed on the application of appropriate statistical methodology and clear interpretation of results in the context of the research question of interest. Support materials for teachers and students have been developed including online resources funded by the Nuffield Foundation's Q-Step programme.

Keywords

Teaching; Teaching statistics; Statistical literacy; Statistical software;

1. Introduction

Many organisations are struggling to recruit staff with skills in data science. It is thought that Scotland needs about 13,000 additional workers each year to cope with the demand for data skills [1]. Two of Scotland's universities have launched a drive to train 100,000 Scots in data skills over the next decade in an effort to transform the country into the 'data capital of Europe'.

The Royal Statistical Society (RSS) has called on the UK government to recognise the value of data and the importance of having the capability to use it [2]. The growth in data collected by technology will play a huge part in the economy of our country and in our personal lives. Investment in education and training to support industry growth is the only way to tackle the gap in skills and preparation to make use of the information stored in data needs to start in early school.

The Making Maths Count Group was set up by the Scottish Government to consider ways of improving confidence and fluency in maths and numeracy for children, young people, parents and those who deliver maths education,

to raise attainment and achievement across learning. The Group published a report in September 2016 containing ten recommendations [3]. Recommendation number 5 states that 'Education Scotland should evaluate the quality of children's and young people's learning in attainment in maths and share examples of good practice.' Around forty secondary, primary and early learning and childcare settings are currently being inspected in response to this recommendation and a final report is due to be published in August 2019. Recommendation number 8 focuses upon improving maths skills for employment including understanding data and statistics.

Education in school forms the foundation for further education and training. Employers have an unmet demand for people with strong literacy and numeracy skills that can be applied in a variety of problem solving contexts, and there is a substantial skills gap in data-driven industries. Work in the data economy is strengthened by successful training in statistics and data analytics combined with skills in critical thinking and the communication of findings. Training and professional development for teachers will be a key ingredient for success. In response to meeting these challenges, 'Developing the Young Workforce' (DYW) is a seven year programme that aims to better prepare children and young people for the world of work. Working with partners there is now an expectation that all schools provide flexible pathways meeting the needs of all learners and ensuring positive destinations for all young people. The Higher Statistics Award has been designed to support the aims of this programme by providing a stand-alone qualification at Higher level that can be delivered by staff working in a variety of curricular areas, thereby offering young people an alternative curricular pathway and providing skills for learning, life and work.

The aim of teaching statistics at school level is to develop statistically literate young adults who make reasonable assumptions when presented with a set of data. Too often it is the case that while students develop adequate procedural knowledge of how to compute descriptive statistics and construct graphs, they are less able to reason about where and when to use these [4]. The RSS has stated 'We would like to see basic numeracy and statistical literacy included in teacher training and in continuing professional development (CPD) for the whole profession' [5]. They see this training of teachers from primary school through to university lecturers as a key ingredient for success in encouraging data literacy in young people. They further emphasise the need to ensure that the training is focussed less on mathematical theory and more on the application of practical skills to real-life situations. CPD training for teachers should be applicable since management of data is a major part of their working lives. [6, 7].

This paper describes the development of the Scottish Qualifications Authority (SQA) Statistics Award in an effort to address these needs.

2. Methodology

The SQA Higher Statistics Award [8] was the result of meetings with Deans of Science and Engineering in Scotland which highlighted a lack of data analysis and problem solving skills in higher education. Further consultations were done with several Higher Education institutions to ensure the course content and learning outcomes would be appropriate and relevant and a working group was tasked to generate ideas.

Meetings with various stakeholders i.e. schools, colleges, universities and employers, helped to identify the key statistical skills to be included in the course curriculum. It was clear that the course should be light on mathematical theory with the emphasis being on the application of data science skills. In addition, the course should be available to pupils of all ability levels across the school curriculum.

Course content was agreed and resources and materials designed. These included support notes, course specification and assessments, teaching and learning resources, access to real life data and computer software training and support. Once developed, the award was promoted at CFE implementation events and CPD training for teaching staff was offered at the Universities of Strathclyde and Edinburgh. These events highlighted a need for additional resources, training and support which lead to further development and revision of the course.

3. Results

The overall aim of the Higher Statistics Award is to develop knowledge, skills and understanding in statistical methods and techniques applied to a variety of real-life contexts from across the curriculum, some of which may be new to the learner. Candidates who complete this qualification will be able to:

- use and apply statistical skills in real-life contexts
- identify and perform an appropriate statistical analysis on given data set(s) using a statistical software package
- communicate the results of a statistical analysis, clearly and concisely, in the context of the problem being addressed

3.1 Course content

Content was devised in collaboration with industry and education to ensure fundamental data science skills covering a wide range of applications would be covered within the course. There are three main topics: an introduction to statistics, correlation and regression and hypothesis testing.

1. Introduction to Statistics: includes types of data; summarising data graphically and numerically using appropriate descriptive statistics; sampling and data distributions (specifically the normal distribution)

2. Correlation and Regression: the simple mathematics of a correlation coefficient is described in detail to ensure a good understanding of the concept and specifically when the use of correlation is inappropriate; mathematics underlying regression modelling is covered lightly, with no need of an understanding of differentiation; model predictions and prediction intervals are explained and goodness-of-fit is assessed using the coefficient of determination; passing mention is made of multiple regression and analysis of residuals
3. Hypothesis Testing: basic ideas; null and alternative hypotheses; p-values; Type 1 and Type 2 errors; confidence intervals; study design; tests for paired and independent samples

3.2 Approaches to learning and teaching

Effective learning and teaching draws on a variety of approaches to enrich the experience of learners. In the Statistics Award, a mix of approaches are used which provide opportunities for personalisation and choice to help motivate and challenge learners. Some of these approaches include: interdisciplinary learning, cross-curricular approaches, investigative and problem solving approaches and resource based learning and e-learning.

To achieve this aim, learners will engage in a variety of learning approaches and activities as appropriate to the understanding of statistical concepts, for example:

- using active and open-ended learning activities such as scientific research, case studies, project-based tasks and presenting findings to others.
- using real-life contexts and experiences that are familiar, unfamiliar and relevant to young people, to meaningfully hone and exemplify skills, knowledge and understanding
- making use of the internet to draw conclusions about specific issues
- recording, in a systematic way, the results of an investigation from different sources
- participating in group work with peers and using collaborative learning opportunities to develop team-working skills
- develop problem solving and critical thinking
- use of questioning and discussion to engage learners in explaining their thinking and check their understanding of fundamental concepts
- making links in themes which cut across the curriculum to encourage transferability of skills, knowledge and understanding — including technology, geography, sciences, social subjects, mathematics, applications of mathematics and health and wellbeing
- using written and/or oral communication and presentation skills to present information

- using appropriate technological resources (e.g. web-based resources)
- using appropriate media resources (e.g. video clips)

Some of these learning and teaching activities may be carried out on a group basis and, where this applies, learners can also receive feedback from their peers. Where possible, teachers and lecturers should provide opportunities to personalise learning and enable learners to have choices in approaches to learning and teaching. There may be opportunities to contextualise approaches through mini-projects or case studies.

Teachers and lecturers should also create opportunities for, and use inclusive approaches to learning and teaching. This can be achieved by encouraging the use of a variety of strategies suited to the needs of all learners. Innovative and creative ways of using technology are valuable in creating inclusive learning and teaching approaches.

3.3 Use of Statistical Software

In order for learners to be equipped with skills in practical data analysis, it was widely agreed that use of a statistical software package should be a key part of the course and, for ease of use, Minitab [9] was the initial software of choice. The introduction of a practical element to the course ensures that candidates are skilled in applying statistical methods to real world problems and able to manipulate and analyse large data sets. The use of a propriety statistical package within the teaching framework also sets this course apart from the Advanced Higher Statistics course offered by the SQA which is more focussed on probability theory and the mathematical aspects of statistics and does not have a practical data analysis component [10].

The key learning outcomes require candidates to be proficient in using a statistical software package for data manipulation and analysis. The licence cost associated with Minitab proved to be prohibitive for schools with limited budgets and R Studio [11] was considered for the practical part of the course. The main advantage of R Studio is that it is a free, open source statistical analysis package and therefore makes the course accessible to all schools in Scotland. The disadvantage is the lack of ease of use. The use of R Studio for the statistics unit was piloted on a group of local school children, aged 15-17, at the University of Strathclyde. The candidates were already using Minitab and subsequently used R Studio to tackle the same research questions. There was no clear preference among those candidates for Minitab over R. In order to ensure that the use of R Studio would not be a barrier to learning, and also to help in training for teaching staff, an online resource was developed for full implementation of the course using R Studio. The development of this resource was funded by Q-Step and two interns – both undergraduate psychology students – produced a web-site with complete teaching and learning resources using R Studio [12].

3.4 CPD Provision

Since the RSS has advocated CPD for teaching staff so strongly, it was important to provide training and support in teaching the material and in the use of R Studio. Several members of the teaching profession had tried to use R themselves with little success, mainly due to time constraints and accessing relevant learning materials. CPD events were again provided for teaching staff on basic R skills using the online resource produced by the Q-Step interns. This proved to be a successful way of training teachers to deliver the course with confidence and enthusiasm. Some of the positive feedback from the CPD training sessions included the following:

- *'I was an attendee at the Higher Statistics event last week. I would just like to thank you in terms of the delivery of unit information etc. Your knowledge and experience oozed and it was a very enjoyable day. Thank you.'*
- *'I particularly appreciated the way you gave real life experiences to back up statements and statistical theories.'*
- *'The time in the labs was very productive. I hadn't used Minitab since I was at University and it had certainly changed from the Windows Manager 3.1 version. Also changed days from looking up t tests and F tests values from books, that I remember.'*
- *'RStudio, I had never used before so to have hands on experience was very worthwhile. In terms of no cost this would be the one we would choose (for obvious reasons).'*
- *'I just wanted to thank you for the presentation on Thursday. I found it incredibly insightful and interesting. The demonstration on using R studio with be invaluable, as I was previously using Excel, which was unnecessarily cumbersome for some things.'*
- *'Going forward I will be "selling" the Stats Award to pupils as an invaluable set of skills for those planning on doing any course at university with a research element.'*
- *'Yesterday's course was very enjoyable. Thank you very much. I came away from yesterday's course very excited about doing statistics. I would like to consider doing more work on statistics at university level, what would be the best pathway for this? Any information would be appreciated.'*

3.5 Assessment

The Higher Statistics Award is assessed in a variety of ways and can include tasks and activities, for example:

- elements of practical assignments such as a project or investigation
- specific assessment tasks or activities
- discrete tests or question papers

These approaches can be used to provide a varied and integrated assessment experience and to make it more coherent and meaningful for learners. It can include learners being tasked to analyse a set of data using a variety of techniques, to answer questions from non-statistical users, and to present the analysis and conclusions in verbal or written form to address the aim of the investigation.

4. Discussion and Conclusion

The SQA Statistics Award introduces pupils in Scotland to a basic knowledge of statistics. More significantly, it allows them to develop skills in applying that knowledge to solve real life questions by interrogating data. The use of propriety statistical software packages like Minitab and R Studio allows candidates from an early age to start developing the computer literacy required to analyse and make sense of the vast amount of data which is available across all aspects of life from health-care to retail. The approaches to learning and teaching this unit promote skills in decision making, communication and critical appraisal of information which are key skills for learning, life and work. In addition, it provides a formal qualification for learners in these skills.

Presently the SQA are looking to incorporate the Statistics Award into a higher level qualification on Applications of Mathematics providing an important curricular pathway for young people. This would make the course more accessible to students across the country and is likely to be a useful mathematical qualification for students going on to further education in non-STEM subjects. This brings Scotland in line with other countries where statistics is taught at primary and secondary education level including New Zealand [13] and the United States [14].

Acknowledgements

The authors would like to sincerely thank the Scottish Qualification Authority for commissioning this work in an effort to address the gap in the school curriculum for education in data science and for funding CPD courses for Scottish teachers to deliver the award. We would also like to thank the teaching staff and pupils who provided invaluable feedback during the development phase of this work and their assistance in the choice of statistical software package. We are also grateful to Emmi de Vries and Kirsty English who were funded by Q-Step to develop the online teaching videos and resources for R Studio.

References

1. Drive launched to train 100,000 Scots in data skills, BBC News, September 2018. Available from <https://www.bbc.co.uk/news/uk-scotland-scotland-business-45629533>.
2. Royal Statistical Society, Education and skills for the data economy. Available from http://www.rss.org.uk/RSS/Influencing_Change/Data_manifesto/Education_and_skills_for_the_data_economy/RSS/Influencing_Change/Data_democracy_sub/Education_and_skills_for_the_data_economy.aspx?hkey=9cd80f64-6260-4502-97cf-6c1f47af2a72.
3. Transforming Scotland into a maths positive nation: final report of the Making Maths Count group. Available from <https://www.gov.scot/publications/transforming-scotland-maths-positive-nation-final-report-making-maths-count/>.
4. Leavy, A. and Hourigan, M. Crime scenes and mystery players! Using driving questions to support the development of statistical literacy. *Teaching Statistics*, 38, 1, p29-35.
5. Education and skills for the data economy. Available from http://www.rss.org.uk/RSS/Influencing_Change/Data_manifesto/Education_and_skills_for_the_data_economy/RSS/Influencing_Change/Data_democracy_sub/Education_and_skills_for_the_data_economy.aspx?hkey=9cd80f64-6260-4502-97cf-6c1f47af2a72#endnotes.
6. Independent Teacher Workload Review Group (2016) Eliminating unnecessary workload associated with data management (PDF). Available from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/511258/Eliminating-unnecessary-workload-associated-with-data-management.pdf.
7. Stokes, M. (2016) Making the most of data in schools (PDF), London: Nesta. Available from <http://www.nesta.org.uk/sites/default/files/making-the-most-of-data-in-schools.pdf>.
8. SQA Higher Statistics Award. Available from <http://www.sqa.org.uk/sqa/73224.html>.
9. Minitab Inc. Available from www.minitab.com.
10. SQA Advanced Higher Statistics. Available from <https://www.sqa.org.uk/sqa/48506.html>.
11. R Studio. Available from <https://www.rstudio.com/>
12. Stats Award. Available from <https://statsaward.weebly.com/>

13. Wild, C.J.; Pfannkuch, M.; Regan, M.; Horton, N.J. (2011). Towards more accessible conceptions of statistical inference. *Journal of the Royal Statistical Society, Series A*. 174 (2): 247– 295.
14. Lindsay, Bruce G.; Kettenring, Jon; Siegmund, David O. (August 2004). A Report on the Future of Statistics. *Statistical Science*. 19 (3): 387–407 (page 403).